Cue-Invariant Geometric Structure of the Population Codes in Macaque V1 and V2

Corentin Massot [*†1,2,3], **Xiaoqi Zhang** [*4,5], **Zitong Wang** [*6,7], **Harold Rockwell** [8], **George Papandreou** [‡9], **Alan Yuille** [10,11], **and Tai-Sing Lee** [†1,2,12]

[1]Computer Science Department, Carnegie Mellon University, Pittsburgh, USA

[2]Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, USA

[3]Department of Neurobiology, University of Pittsburgh, Pittsburgh, USA

[4]Department of Computational Biology, Carnegie Mellon University, Pittsburgh, USA

[5]Department of Biological Sciences, Purdue University, West Lafayette, USA

[6]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, USA

[7]Department of Ophthalmology, University of Pittsburgh, Pittsburgh, USA

[8]Committee on Computational Neuroscience, University of Chicago, Chicago, USA

[9]Department of Statistics, University of California, Los Angeles

[10]Department of Cognitive Science, Johns Hopkins University, Baltimore, USA

[11]Department of Computer Science, Johns Hopkins University, Baltimore, USA

[12]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, USA

December 5, 2023

ABSTRACT

We investigated the cue-invariant representation of visual patterns associated with surface boundaries in V1 and V2. We found that individual neurons exhibited a modest degree of tuning invariance in their responses to these patterns. This tuning invariance is stronger in V1 than in V2. At a population activity level, we studied the performance of a decoder trained with one cue in decoding patterns defined in another cue. We found that cue-transfer decoding is greatly enhanced when a geometric transform is first performed to align the population activities across cues. With this geometric transform, transfer decoding can be successful even when the tuning invariance of the individual neurons are destroyed by shuffling and when the neurons are from distinct populations. These findings suggest that abstract representation of these boundary-related patterns in V1 and

[*]Equal contribution

[†]Corresponding authors: `corentinmassot@pitt.edu, taislee@andrew.cmu.edu`

[‡]George Papandreou contributed to this project while employed at UCLA. He is currently working at Snap Inc., London, United Kingdom.

V2 might be primarily encoded in the geometric structures of the population codes, rather than the cue-invariant tuning properties of the individual neurons.

We can perceive objects and scenes depicted in line drawings, sketches, or cartoons as effortlessly as we can perceive them in photographs. This is an example of the remarkable capacity of the visual system to recover the same information from differing renderings of objects and scenes. This perceptual ability appears to be innate as it has been observed in humans such as human babies (Yonas and Arterberry, 1994) and stone-edge tribe members with no experience with image-based communication tools (Sayim and Cavanagh, 2011), as well as in animals, such as chimpanzees (Tanaka, 2007). Despite the disparate physical cues used in these different types of rendering, they often evoke a consistent and unambiguous perceptual interpretation, similar to that of photographs. From a neuroscientific point of view, this perceptual capacity can be interpreted as manifestations of a common underlying neural representation of an abstract perceptual construct of objects in natural scenes. What is the neural basis of these abstract representations in the visual cortex?

To answer this question, we investigated the representational invariance capacity of early visual areas (V1 and V2) to different rendering of local patterns associated with fragments of surface boundary discontinuity in natural images. V1 has been shown to encode oriented segments (Hubel and Wiesel, 1968) but also more complex shapes(Gallant et al., 1993; McManus et al., 2011; Tang et al., 2018). V2 has been shown to encode complex shape elements and composition of orientation features (Anzai et al., 2007; Hegdé and Van Essen, 2000, 2007) and be sensitive to texture patterns (Freeman et al., 2013; Laskar et al., 2020; Ziemba et al., 2016, 2018, 2019). Here, we focused on the representation of a set of local surface boundary features that are extracted from natural scenes. They are called boundary visual concepts. Abstract representation of visual concepts is fundamental to the compositional theory of vision (Geman, Yuille), which dictates that visual objects are represented structurally in a compositional hierarchy. Elementary concepts are represented in early visual areas, and more global and complex structures are constructed downstream using these elementary structures in a compositional manner, ultimately leading to object representation. Compositional structures allow modules to be dynamically reconfigured to provide a flexible representation of objects. Boundary visual concepts are elementary features used in compositional systems in computer vision (Yuille), extracted from the boundary segmentation dataset using clustering techniques (see Supplementary Information). They include oriented lines, curves, junctions of varying complexity. Our study is designed to understand better how abstract representations of these boundary visual concepts are represented in the brain. Each of the visual concepts to be tested is rendered in the form of either line drawings (Edge Contour EC), cartoons (Appearance Contour AC), and photographs (Examplar Contour EX).

Along the ventral visual pathways, early visual areas such as V1 and V2 might provide the most appropriate infrastructure for representing these boundary elements because of the precision of their representation in both features and spatial dimensions as well as their recurrent connections known for mediating boundary computation (Fitzpatrick, 1996; Gilbert et al., 2000; Lee and Mumford, 2003; Yan et al., 2014). Using linear-multiple-electrode arrays in two awake-behaving macaque monkeys, we recorded the response of V1 and V2 neurons to local boundary concepts. We studied the correlation between the tuning curves of individual neurons in response to these local visual concepts represented in the different cues and examined where tuning correlations might be the strongest along the different laminae of V1 and V2 to track the development of cue invariance. To assess cue invariance at the population level, we developed a metric

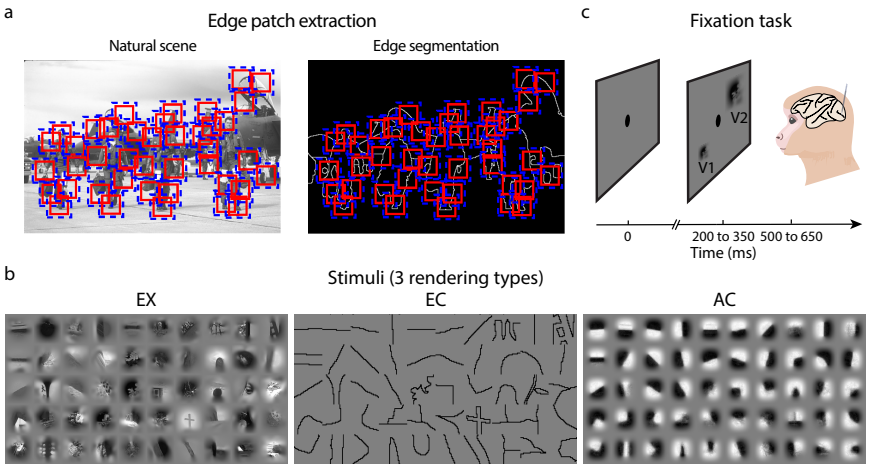

Figure 1: Visual stimuli and experimental setup. (a) Patches containing elements of boundary surface such as edges, contours or abrupt changes were extracted from natural scenes; left, extracted patches containing the original information from the natural scene; right, extracted patches from natural scenes after application of an edge segmentation technique to recover only the edge information. (b) Complete set of visual stimuli divided into 3 rendering types EX (left), EC (middle) and AC (right). (c) Experimental setup: stimuli were displayed on a computer screen; after presentation of a fixation dot, indicating the start of a new trial, a stimuli was presented after a delay; the animal had to maintain fixation during the presentation of the stimuli until its disappearance; the size of the stimuli were adapted to the average size of the receptive fields in V1 (smaller stimuli) and V2 (larger stimuli); a laminar probe was acutely inserted in V1 or V2 to record neural activity during the task.

based on cue-transfer decoding, which uses a linear decoder trained on the neural responses to visual concepts rendered in one cue to decode the concepts rendered in a different cue. We found that the geometric structure of the population codes, rather than individual neurons' tuning correlation, might be key to cue-invariant representation of these boundary visual concepts and that these geometric structures are established earlier on in the primary visual cortex.

Results

We recorded the activity of 157 V1 neurons (48 for monkey FR and 109 for monkey KO) and 258 V2 neurons (112 for monkey FR and 146 for monkey KO) in two non-human primates using linear multi-contact electrodes (see Methods for more details). We analysed the responses of V1 and V2 neurons to the presentation of boundary concept stimuli with varying shape complexity reflecting the natural statistics of the visual environment (figure 1.a). To compare invariance properties between the two regions, we adapted the size of the stimuli to the size of the receptive fields (#° and #° for V1 and V2 neurons respectively (figure 1.c). Figure 1.b shows examples of the three sets of stimuli that were used during the experiment. There are 50 Edge Concepts (EC) stimuli extracted as 50 cluster means of local patch of surface segmentation boundaries from the Berkeley segmentation database. These cluster means are prototypes of boundary segments of varying shape complexity. They resemble fragments of line drawings. Appearance Concept stimuli (AC) represent the first principal component of the image patches associated with each of the 50 Edge Concepts. They resemble patches of cartoon representation of appearance associated with the 50 edge concepts. Exampler stimuli

(EX) are the original patch extracted from the photographs of natural scenes and containing a surface boundary. These stimuli displayed a variety of texture and complex patterns defining and surrounding each of the 50 edge concept (see Methods and Supplementary Material for details on the design of the stimuli dataset). As a result, we obtained 3 dataset made of 50 stimuli each representing the same surface boundaries but rendered differently.

Correlation of Individual V1 and V2 neurons to Boundary Concepts across Cues

First, we examined at invariance coding at the individual neuron level. To evaluate the cue invariance properties of visual neurons in V1 and V2, we measured the response of neurons to the three sets of boundary concept stimuli and computed the correlation of each neuron's tuning to the boundary concept stimuli between two cues. Figure 2.a shows the three set of stimuli rank-ordered by the response of an example V1 neuron recorded in monkey KO. Through visual inspection, one can observe that this neuron displays preference to the same surface boundary across the different rendering type (for example compare the first and last rows across the three sets). When plotting the response of this neuron to each rank-ordered set of stimuli, we obtain tuning curves that are non-overlapping (figure 2.b). So a direct comparison based solely on the response firing rate between the three rendering types is not appropriate. Instead we analyzed the correlation between each pair of tuning curves. Figure 2.c shows that the tuning correlations for the example V1 neuron for the all three pairs of rendering type (EC-AC, EC-EX and AC-EX) are significant (Pearson correlation, $p < 0.01$ for all three correlations). This example V1 neuron shows some degree of invariance across the three rendering types (i.e. significant tuning correlation for the triple pairs of stimuli). We quantified the proportion of neurons in V1 and V2 that showed significant tuning correlations for single, double and triple pairs of stimuli (see figure 2 caption for the full list of correlation sets). Figure 2.d shows the proportion of neurons with significant tuning correlations on the whole population of V1 and V2 neurons recorded in each monkey. For monkey KO, the proportion of V1 neurons showing invariance was significantly above chance for single (on average, 34% of the total number of V1 neurons), double (25%) and triple (22%) pairs of rendering types (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). The proportion of V2 neurons showing invariance was reduced compare to V1 but was still significantly above chance for single (on average, 14% of the total number of V2 neurons), double (6%) and triple (5%) pairs of rendering types. The proportion of neurons was significantly higher in V1 than in V2 for all correlation sets (comparison of population proportions, $p < 0.0.1$). For monkey FR, the results were similar albeit the proportion of V1 and V2 neurons showing cue-invariance was reduced compare to monkey KO. For V1 neurons, it was also significantly above chance for single (22%), double (10%) and triple (6%) pairs of rendering types. The proportion of V2 neurons showing invariance was also reduced compare to V1 but was still significantly above chance for single (13%), double (5%) and triple (4%) pairs of rendering types. The proportion of neurons was also significantly higher in V1 than in V2 for all correlation sets (comparison of population proportions, $p < 0.0.1$).

Next, to gain a deeper understanding of the cortical processing underlying cue-invariance in V1 and V2 we analyzed neural tuning invariance for different cortical layers. An estimation of the laminar positions of the neurons within V1 and V2 was rendered possible by the use of a multi-contact laminar probe. We used a combination of CSD analysis and onset latency (figure s2, see figure s2 and Materials and Methods) to estimate the laminar position of neurons and group them into three main layer ensembles, namely 'Superficial', 'Layer 4' and 'Deep' layers. Figure 2.e shows individual V1 neurons' tuning correlations for both monkeys grouped by their cortical layer. For both monkey KO and FR, one salient result is a higher proportion of V1 neurons within layer 4 showing invariance compare to the other

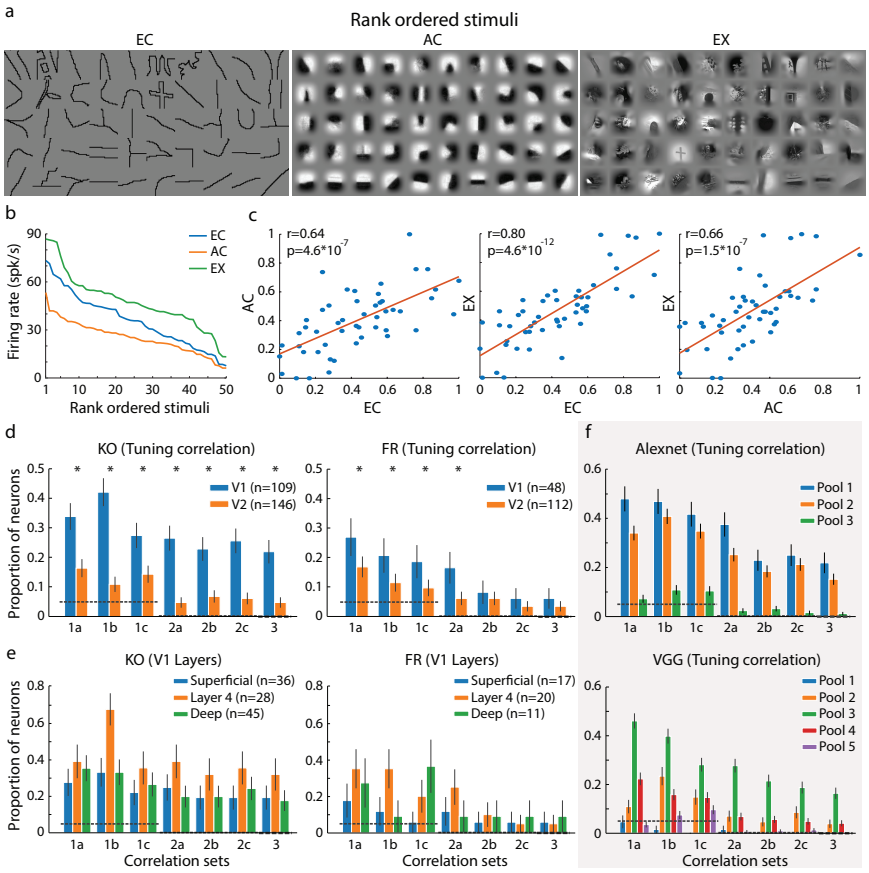

Figure 2: Tuning correlation of individual neurons in V1 and V2. (a-b-c) Tuning correlation results of an example V1 neuron in monkey KO; (a) Rank-ordered stimuli of each rendering type based on the response of the example neuron. (b) Tuning curves of the example neuron for each rendering type. (c) Scatter plots showing the correlation between pairs of tuning curves (red line); insets display the Pearson correlation and associated P-value. (d) Tuning correlation results for V1 (blue) and V2 (orange) neurons for monkey KO (left) and FR (right) for the different correlation sets; a correlation set correspond to single, double or triple pairs of stimuli sets (belonging to two different rendering types); Correlation sets are coded as follows: EC-AC (1a); EC-EX (1b); AC-EX (1c); EC-AC and EC-EX (2a); EC-AC and AC-EX (2b); EC-EX and AC-EX (2c); EC-AC and AC-EX and EC-EX (3); bars indicate the proportion of neurons that displayed significant tuning correlations for each correlation set; error bars indicate confidence intervals using SSE; horizontal dashed lines indicate chance level for single (P=0.05), double (P=2.5×10^{-3}) and triple (P=1.25×10^{-4}) tuning correlation pairs; insets indicate the number of neurons in each region; * symbol at the top of the plot indicates a significant difference between V1 and V2 proportion of neurons using a comparison of population proportions. (e) Tuning correlation results for V1 neurons in superficial (blue), layer 4 (orange) and deep (green) layers for monkey KO (left) and FR (right) for the different correlation sets; same display as panel d; insets indicate the number of neurons in each layer. (f) Tuning correlation of artificial units from the pooling layers of AlexNet (top, 3 pooling layers) and VGG (bottom, 5 pooling layers); same display as panel d.

layers. For monkey KO, the average proportion of V1 neurons withing Layer 4 was higher for all correlation sets, but only significant for EC-EX (1b) correlation pair (on average, 67% of the total number of V1 neurons, comparison of population proportion, P=3.5×10^{-4} (layer 4 vs. superficial layers) and P=7.8×10^{-4} (layer 4 vs. deep layers)). For monkey FR, the average proportion of V1 neurons withing Layer 4 was also higher for several correlation sets (EC-AC (1a), EC-EX (1b) and EC-AC and EC-EX (2a)), and, like monkey KO, it was significant for EC-EX (1b) correlation pair (on average, 35% of the total number of V1 neurons, comparison of population proportion, P=2.5×10^{-3} (layer 4 vs. superficial layers) and P=9.2×10^{-4} (layer 4 vs. deep layers)). Some additional observations can be made for monkey FR about superficial (lack of significant proportion of cue-invariant V1 neurons) and deep (higher or equal average proportion of neurons compared to layer 4), however the same patterns were not observed in monkey KO. The proportion of cue-invariant V2 neurons for both monkeys grouped by their cortical layer (Figure s3.a) did not show any consistent pattern.

Finally, to investigate whether the cue-invariant properties of individual neurons depended on the early integration time of the neural activity or from a putative late feedback from downstream regions, we re-analyzed the neural tuning invariance distinguishing two temporal integration windows (early (40-110ms) and late (110-180ms) windows). Figure s3.b-c shows the proportion of cue-invariance neurons for early and late temporal windows for both monkeys in V1 and V2. For both monkeys, individual V1 neurons showed significant tuning correlation across all correlation pairs whether the neural activity is analyzed within the early or the late window (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). The proportion of V1 neurons showing significant cue-invariance was, on average, very similar across temporal windows: 30%, 20% and 16% (resp. 28%, 19% and 17%) of the total number of V1 neurons for single, double and triple pairs of rendering types for monkey KO for activity measured within the early (resp. late) window; 21%, 10% and 6% (resp. 18%, 6% and 4%) of the total number of V1 neurons for single, double and triple pairs of rendering types for monkey FR for activity measured within the early (resp. late) window. These results allow us to conclude that, in V1, the proportion of cue-invariant neurons did not depend on the window of analysis.

In V2, the proportion of individual neurons showing tuning correlation in the early window was significantly reduced compared to V1 (comparison of population proportions, $p < 0.01$) for both monkeys. In addition, for monkey KO, the proportion of V2 neurons showing significant tuning correlation was significant across all correlation pairs (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). For monkey FR, this proportion was largely reduced and only reached significant for EC-EX (1b) and EC-AC and EC-EX (2a) correlation pairs (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). This pattern of results in the late windows of V2 neurons was somewhat reversed. The proportion of individual neurons showing tuning correlation in the late window was significantly reduced compared to V1 (comparison of population proportions, $p < 0.01$) only for monkey KO. For monkey KO, the proportion of V2 neurons showing significant tuning correlation was largely reduced across all correlation pairs and did not reach significance for EC-EX (1b) and AC-EX (1c) correlation pairs (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). For monkey FR, this proportion was on the contrary still large and significant across all correlation pairs (significance was reached when the minimum of the Wilson score interval was above chance level of the correlation set). These contradictory results prevent any conclusion about the dependence of the proportion of cue-invariant neurons in V2 on the window of analysis.

Overall, at individual neuronal level, more neurons were cue-invariant in V1 than in V2, and, in particular a significant proportion of V1 neurons showed cue-invariance across multiple pairs of rendering type. The proportion of neurons showing invariance was higher in monkey KO than in monkey FR, which may be due to differences in their visual experience and training. These findings suggest that early visual areas such as V1 may represent basic visual concepts, such as oriented edges, in a more cue-invariant manner. in the following, we analyzed how cue-invariance was reflected at the population level.

Invariant Geometric Structures in V1 and V2 Population Codes

Next, we evaluated cue-invariant representation in the population responses to these boundary concepts, first by the standard Representation Similarity Measures that measure the consistency in neural codes between a pair of neurons across cues, and then by cue-transfer decoding, which is designed to assess the consistency in the geometry of the constellation of population codes to the different boundary concepts.

We first evaluated the invariance properties of the neural response in V1 and V2 at the population level using representation similarity between responses to the rendering types. Representation similarity analysis (RSA) is a useful method to capture pattern information a brain region conveys about the experimental task ((Kriegeskorte and Kievit, 2013)). It gives a description of the information content of the population activity that does not depend on assumptions about the decoding mechanism nor make any assumption about the geometry of the population code. Here for each visual region and each rendering type, we generated the similarity matrix of the response of each neuron to one rendering type and measured its correlation with the similarity matrix to a second rendering type. A significant amount of correlation indicates some degree of invariance is maintained between the two population responses. Figure s6.a shows the result of the RSA analysis for all pairs of rendering type of the population response in V1 and V2. In both monkeys, we found that the similarity is significant for EC-AC and EC-EX pairs in V1 and V2. For the AC-EX pair, only V1 population response in monkey KO reached significance. Notice that the upper bounds on the similarity measurements (see Materials and Methods) are below 1 and reflect some intrinsic noise of the population activity. If we regularize the RSA measures by their upper bounds, the results stay the same. RSA shows that there is a significant level of representational similarity between EC-AC and EC-EX pairs of rendering types for both monkeys but not for the AC-EX pair. This is understandable because Edge Concepts (EC stimuli) are the anchoring prototype stimuli AC and EX stimuli are based on. There is no direct relationship between AC and EX stimuli.

RSA analysis provides information about the consistency between the neural codes between pairs of stimuli but it does not provide information about the geometry of the constellation of population codes for the entire set of stimuli. We propose a new approach, called cue-transfer decoding, to assess the invariance of the geometry of the neural codes. Cue-transfer decoding uses a linear decoder trained on the neural responses to visual concepts rendered in one cue to decode the concepts rendered in a different cue. Its performance relies on the consistency in the global configuration of the population codes to the different boundary concepts across cues. A naive application of cue-transfer decoding to the population codes, with the mean population codes of all stimuli of one type of rendering subtracted, yields a performance that is two or three times better than chance (2 %). When we projected the population codes in their first two principal components, we noticed that the geometry of the constellation of the population codes of the boundary concepts rendered in one cue bears some similar similarity to the geometry of the population codes of the boundary

concepts rendered in another cue. We applied Procrustes Transform (PT) to align the population codes in one cue to the population codes in the other cue by rotation, translation and scaling so that the sum of the Euclidean distance between the neural representations of all the corresponding boundary visual concepts is minimized. Figure 3a shows an example of Procrustes transformation applied to the population responses of V1 neurons to the three rendering types in the space of the first two principal components. Procrustes transformation was applied to the population responses of EC and AC rendering type against the population response to EX stimuli. After applying PT, one can observe that the projections of the three sets of population codes exhibit a greater degree of geometric correspondence. After alignment with PT, we found cue-transfer decoding performance yields significantly better performance. We call the naive cue-transfer decoding non-PT decoding and the cue-transfer decoding with Procrustes Transform PT decoding.

Figure 3.b compared the performance of PT decoding with non-PT decoding for V1 neurons and V2 neurons. For 50 stimuli, the chance decoding accuracy is 2%, indicated by a horizontal dash-line on the bars. We found PT cue-transfer decoding is significantly better than non-PT decoding, and then cue-tranfer decoding is significantly better for V1 than V2. Better performance means the geometry of the population code is more invariant across cues. The better performance of V1 over V2 is in part due to the fact that V1 population code is more discriminative of the boundary visual concepts in each rendering than V2, as shown in the self-coding in which a classifier trained to decode the 50 stimuli rendered in one cue was used to decode stimuli in the hold-out trials in the same cue. Thus, this performance provided the upper bound of the decoding performance of the neurons due to trial-by-trial variability and tuning properties. Figure 3b shows that non-PT transfer decoding is significantly better than chance, achieving about 15% of the achievable decoding accuracy established by the upper bound set by self-decoding. However, with Procrustes transform, which allows the constellation of neural population vectors for this set of stimuli to be rotated, translated and stretched, transfer decoding can achieve 30-40

The correlation of the individual neurons' tuning curves in the different cues contributes to the performance of non-PT transfer decoding. On the other hand, PT transfer decoding measures the similarity in the geometric structures of the neural representation of 50 stimulus patterns in different cues at the population level and might not depend on the tuning correlation of individual neurons. To evaluate whether the tuning correlation of the neurons matter or not in cue-transfer decoding, we performed a shuffling experiment in which the identity labels are shuffled among the neurons, but the stimulus and response pairs are preserved. For example, while for the data for one cue was kept the same, the neurons' identities are shuffled randomly in the other cue. Shuffling completely destroyed the tuning correlation between a neuron across the different cues (Figure 1). With shuffling, we found that non-PT transfer decoding was reduced to chance, as shown in Figure 3d, but PT-transfer decoding performance remains intact, suggesting PT decoding relies solely on the similarity in the geometric structures of neural representation and the tuning correlation of individual neurons appears to be inconsequential.

We take this argument a step further by performing PT transfer decoding in three other scenarios: (1) classifier based on one cue in one set of V1 neurons is used to decode stimuli based on the same cue in another set of V1 neurons; (2) classifier based on one cue in one set of V2 neurons is used to decode stimuli based on another cue in another set of V2 neurons; (3) classifier based on one cue in a set of V1 neurons is used to decode stimuli based on another cue in a set of V2 neurons and vice versa. We observed that the PT transfer decoding in all these scenarios are no different from those in Figure 3c and 3d, as long as the number of neurons remains the same in the comparisons.

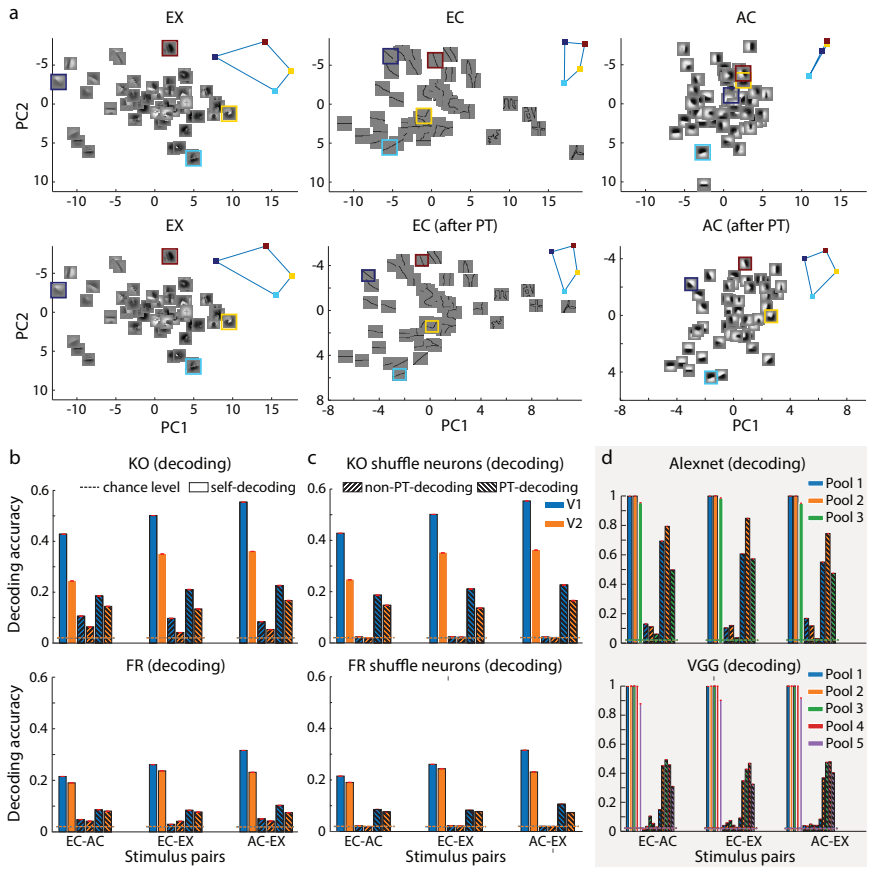

Figure 3: Procrustes transformation and population decoding. (a) Example of the effect of Procrustes transformation on a V1 population for each rendering type; the high-dimension representation of the neural population activity associated to each set of stimuli for the three rendering types was reduced using PCA and stimuli were projected on the 2-dimensional space defined by the two first principal components; first and second row show the representation of the stimuli sets before and after Procrustes transformation; insets show the relative position of four example stimuli (indicated with colored borders): after Procrustes transformation, the geometrical matching of the population activity between the EC and AC with the PCA projection of the EX stimuli has been improved (through spatial rotation and scaling). (b) Population decoding accuracy for monkey KO (top left) and FR (bottom left); the x-axis indicates the stimulus pairs of rendering types used to estimate the decoding performance; blue and red bars indicate V1 and V2 data, respectively; the first pair of bars (plain) indicates the decoding performance within each population of neurons (self-decoding) and represents an upper-bound on the decoding performances; the second pairs of bars (downward hatched pattern) indicates the decoding performance without any transformation (non-PT-decoding); the third pair of bars (upward hatched pattern) indicates the decoding performance after Procrustes transformation (PT-decoding); blue and red horizontal dashed lines indicate chance level for V1 and V2 neural population decoding, respectively; red error bars indicate SEM of the decoding performance. (c) Decoding performance after applying a shuffling procedure on the population of V1 and V2 neurons; same display as panel b. (d) Decoding performance of the population of artificial units from the pooling layers of AlexNet (top, 3 pooling layers) and VGG (bottom, 5 pooling layers); same display as panel b.

These observations suggest that the geometric structures of the neural representations are likely preserved across rendering cues, across distinct neuronal populations within V1 or within V2 and across V1 and V2. Invariant geometric structure in neuronal population code therefore provides a possible robust mechanism for computing and representing cue-invariant representation for perceptual interpretation.

PT transfer decoding does increase with the increase in the number of neurons in the population the set as shown in Figure 4a, as the geometric structure will be approximated better by with a large number of neurons. This suggests the underlying geometric structure of the population codes for the boundary concepts might indeed be cue-invariant in the brain given the number of neurons each hypercolumn is highly overcomplete.

Recent advances in artificial neural networks have led to the development of biology-compatible models of the ventral visual system. The best models are based on convolutional neural networks (CNN), and the analysis of their artificial units displays an increase in their selectivity ((Laskar et al., 2020; Schrimpf et al., 2018)). We applied the same analysis to two convolutional neural networks that have been extensively compared to neural representations in macaque. We found that the intermediate layers of AlexNet and VGG exhibited similar degree of cue-transfer invariance, as shown in Figure 3d, and with enough number of neurons, cue-transfer decoding can reach a fairly high accuracy, suggesting the geometric structures of the population codes for these boundary concepts also exhibit a high degree of cue-invariant in these neural networks (Figure 4b). The neural networks do indicate that at higher layers in the neural networks, but not the highest, the cue-transfer decoding can reach up to 85%.

The success of PT transfer decoding suggests that the invariance of geometric structure of neural representation might be important for cue-invariant interpretation of these boundary concept stimuli. Invariance in the geometric structure of neural representation, when coupled with the Procrustes transform, offers a potential neural mechanism for cue-invariant representation of perceptual interpretation of stimuli. Traditionally, cue-invariance has been investigated at the individual neuron level. This study suggests a different perspective. Perhaps individual neuron's invariant tuning is not critical. As long as there are a sufficient number of neurons to span the stimulus space and capture the geometric shape of the neural representation, a stimulus rendered in two or multiple cues can be brought into alignment into an abstract interpretation downstream by a linear transformation or single-layer perception.

Discussion

Hubel and Wiesel (cite Hube and Wiesel) used bar and contrast edge to probe the orientation selectivity of V1 neurons. They were motivated by the intuition that the neurons might serve as some cue-invariant edge detectors. Similarly, Connors and colleagues (Connor and Pasupathy) also used contours and contrast edge stimuli in their parametric study of the shape and junctions in individual V4 neurons. Studies on more abstract perceptual constructs, such as igure-ground segregation, , have explored cue-invariance in terms of texture, luminance, stereo, and motion (Zipser et al., 1996). These studies focused on the study of the tuning of individual neurons to specific perceptual dimensions and parameters. The assumption is a neuron coding abstract representation should tune to the abstract representation regardless of the visual cues defining it. Our study extended cue-invariant studies in three new directions. First, we studied a set of local visual concepts that are extracted from the segmentation boundaries in natural images. These concepts include straight lines, as well as curves, junctions, and more complex contour patterns. If these contour patterns are legitimate visual concepts or symbolic tokens, neurons should be able to extract and represent them in a

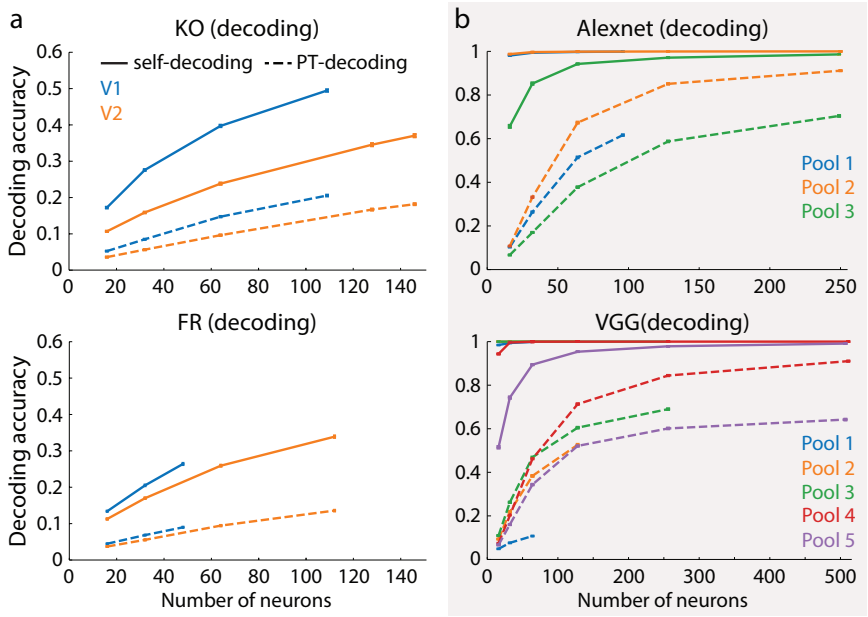

Figure 4: Effect of population size on decoding accuracy. (a) Decoding accuracy of V1 (blue) and V2 (red) data with increasing number of neurons for monkey KO (top) and FR (bottom); solid and dashed lines represent decoding accuracy using the self-decoding and the PT-decoding procedures, respectively; SSE are represented by the horizontal bars centered on the population size used to comput the decoding accuracy (very small). (b) Decoding accuracy of the different pooling layers of AlexNet (top) and VGG (bottom) CNN models; same display as panel a.

cue-invariant manner. To what extent individual neurons' tuning curves to these visual concepts rendered in different cues are correlated? Second, we recorded from different laminar layers of V1 and V2 using linear multi-electrode arrays. Our objective is to track the development of the cue-invariant abstract representation along the ventral visual pathway. Does the tuning invariance across cues increase as the information propagates down the hierarchical visual systems? Third, we investigated how cue invariance could be achieved at the population level. Is cue invariance at the population level a consequence of the tuning correlation of individual neurons? Interestingly, we are surprised to find that the answers to these questions are different from conventional wisdom and common assumptions. Rather, we discover that the geometry of the neural population codes might be key to cue-invariant abstract representation of these visual concepts in the brain.

Encoding of Abstract Boundary Concepts in V1 and V2

Abstract representation of visual concepts is fundamental to the compositional theory of vision (cite Geman, Yuille), which dictates that visual objects are represented structurally in a compositional hierarchy. Elementary concepts are represented in early visual areas, and more global and complex structures are constructed downstream using these elementary structures in a compositional manner, ultimately leading to object representation. Compositional structures allow modules to be dynamically reconfigured to provide a flexible representation of objects. Boundary visual concepts are elementary features used in compositional systems in computer vision (cite Yuille). Our study is designed to understand better how abstract representations of these boundary visual concepts are represented in the brain.

To understand cue invariance at the individual neurons' level, we adopted and generalized the traditional paradigm by computing the tuning correlation of individual neurons to these visual concepts across different cues. We observed (Figure 3) that some neurons exhibited a certain degree of tuning curve invariance, in the sense of statistical significance in the correlation in their tuning curves, for one pair of cues, and some even showed invariance across all three cues. These tuning correlations were not high (what is the distribution?) but statistically significant. It is uncertain whether and how these partial correlations can lead to abstract representation. Surprisingly, we found that a larger proportion of neurons in V1 exhibited significant cue invariance than in V2. In fact, layer 4 of V1, which is the input layer, appears to have a higher proportion of neurons that exhibit statistically significant tuning correlations across cues than the other layers in V1 and V2. This counters our expectation, as we expect cue-invariance to become more pronounced downstream. Yet, upon conducting layer and areal analyses, we found that cue invariance for this set of boundary concepts decreases rather than increases along the hierarchical ventral stream visual pathway. However, our tuning invariance measure is based on the correlation in the tuning curves. When a neuron becomes more invariant (rotation, translation) in the traditional sense, its tuning curve to the stimuli might actually become more flat.The neuron might become less discriminative of the tested stimuli within each cue. The loss in feature and spatial precision due to rotation and translation invarinance might actually reduce the tuning correlation, and hence the cue-invariance of the visual concepts in individual neurons.

It is reasonable to think that each visual area may specialize in extracting a specific set of visual concepts appropriate for the computation in that area. V1 perhaps provides the best representation for computing and representing boundary information, furnishing a high-resolution buffer for geometric computation (Lee and Mumford, 2023). From that perspective, it is reasonable that boundary visual concepts are best represented in V1, even in a cue-invariant manner. The higher percentage of V1 neurons in exhibiting statistically significant tuning correlation in V1 than V2 can be understood in this way. Evidence consistent with this notion also can be found at the population level, where we observe that within-cue decoding (self-decoding in Figure 4) is higher in V1 than V2. That is, controlled for the number of neurons, V1 can discriminate among these sets of stimuli within each cue significantly better than V2.

Cue-Invariant Geometric Structures in Neuronal Population Codes

To assess cue invariance at the population level, we developed a metric based on cue-transfer decoding, which uses a linear decoder trained on the neural responses to visual concepts rendered in one cue to decode the concepts rendered in a different cue. We observed a modest amount of representational similarity in the population codes across cues. However, cue-transfer decoding relies on the consistency in the global configuration of the population codes to the different visual concepts across cues. Representational similarity measures only the pairwise relationship of the neural codes between stimuli. It does not capture the geometric structure of the population code of these concepts. Cue-transfer decoding allows us to assess the invariance or persistence of the geometric structure across cues. However, we found the cue-transfer decoding accuracy on the raw responses of the neuronal population, after simple translation to center the population responses around their mean, provides very weak cue-transfer decoding performance, even though it is statistically significantly better than chance. It is, however, evident in the lower dimensional space spanned by the first two principal components of the population neural responses that the geometry of the population codes of the constellation of visual concepts rendered in different cues appear to be rotated, translated, and scaled versions of one another. Applying Procrustes transform to align the population codes across cues produces significantly better cue-

transfer decoding. This suggests that the geometry of the neural codes is preserved across cues and that the invariance in the geometric structure of the population code might play a far more important role than tuning correlations of individual neurons. In fact, we found that the geometry, measured this way, is preserved when the individual neurons' tuning correlation is completely destroyed by shuffling the neurons' identity across cues. It is also preserved between two populations of neurons within the same area and across V1 and V2. These observations lead us to conclude that the invariant representation of abstract boundary visual concepts could be represented in the geometric structure of the population codes rather than the tuning correlations of the individual neurons across cues. These geometric structures of the population code can easily be read out by one-layer perceptron by a simple mapping between different cues or to a common and abstract representation where line drawings, cartoons, and photographs will share the same perceptual representation.

Materials and Methods

Animal preparation

Two adult Rhesus macaque monkeys (Macaca mulatta) participated in the study (Monkey F and monkey K). Under general anesthesia, each animal was implanted with a head fixation post, scleral search coils, and a recording chamber overlying the occipital operculum, with its center roughly at the border between areas V2 and V1 (more details can be found in (Smith et al., 2007)). All procedures were approved by the Institutional Animal Care and Use Committee of Carnegie Mellon University and were in compliance with the guidelines set forth in the United States Public Health Service Guide for the Care and Use of Laboratory Animals.

Visual task

Eye movements were recorded using a scleral search coil system (Riverbend Instruments, Birmingham, AL), sampled at 200Hz. Behavioral task control was done using NIMH Cortex. After recovery from surgery, each animal was trained to perform a standard fixation task. At the beginning of each trial, a fixation dot appeared at the center of the screen. The animal had 3000 ms to bring its line of sight towards it and maintain an eye position within 0.8° of the central point for 200 to 350 ms. Then, a visual stimulus appeared at a specific location on the screen and the animal was required to maintain its fixation for at least 300ms. If all these conditions were respected, the trial was counted as successful and the animal received a liquid reward. Otherwise, the trial was aborted.

Visual stimuli

The goal of this study was to analyze the response of V1 and V2 neurons to varying edge stimuli from simple to complex shapes reflecting the natural statistics of the visual environment. Our visual stimuli represented edge contour boundaries extracted from natural scenes. Each edge stimulus was displayed using 3 types of rendering: 'Edge Contour' (EC); 'Appearance Contour' (AC) and 'Example Contour' (EX). Here we give the detailed description of their design.

First, we extracted a large number of patches in natural scenes taken from the Berkeley Segmentation Dataset ((Martin et al., 2001)). We used edge annotations done by human experts and used the segmentation binary maps from all images to collect a total of 5818 patches of size 33×33 and 45x45 pixels. Then, we computed pairwise shape distances between

the segmented edges within each patch using Haussdorff shape metric. We clustered the whole set of edges using affinity propagation ((Frey and Dueck, 2007)) into 50 clusters. The assumption was that a relatively small number of exemplars (50) plus their translated versions (up to ±6 pixels in the horizontal or vertical direction) suffices to closely approximate most edge patches occurring in practice. This example-based clustering method allows the cluster centers to be actual examples from the dataset. The centers of each cluster formed the set of 'Edge Contours' (EC) stimuli (Figure 5.c). The original patches extracted from the natural scenes, from which the EC stimuli originated from, formed the set of 'Example' (EX) stimuli (Figure 5.d). Finally, to learn a model of the patch appearance, we performed a principle components analysis (PCA) on the aligned and normalized edge patches of each cluster. The first component was very similar (albeit a reversal of edge luminance polarity) to the mean luminance patch of the cluster. The mean patches of each cluster formed the set of 'Appearance Contour' (AC) stimuli (Figure **??**.e). To prevent border artifacts when measuring V1 and V2 neurons' responses, a circular smoothing aperture was applied to the patch to remove any abrupt edge information. 2 stimuli sizes were used to match the average size of V1 and V2 neurons (33X33 and 64X64 pixels respectively). The selected 50 edge contours and their 3 renderings, resulted in a dataset of 150 stimuli that were used for the visual experiment. Figure 1.b in the main text displays the complete stimuli dataset.

Neurophysiological recordings

We used a 2x12-channel laminar probe (Alpha-Omega, Alpharetta, USA; 150 μm inter-contact distance; 300μm diameter; 1MΩ impedance for each channel) to record neural activity in V1 and in V2. The probe was inserted through the dura into V1 operculum using a hydraulic Microdrive (Narishige, Tokyo, Japan). V1 neural activity was associated to the activity first detected upon the insertion and lowering of the probe. V2 neural activity was recorded by increasing the depth of the probe. The separation of V1 from V2 activity was detected based on a region of clear decreased neural activity (white matter). In each region, the probe was positioned so that at least one contact had a well isolated spike waveform (reference neuron) and neural activity due to visual stimulation was detected on the maximum number of the remaining contacts. Neurophysiological signals were recorded using Blackrock acquisition system (Blackrock, Salt Lake City, USA). Data recording was synchronized with the beginning of the trial, and the timing of all trial events were recorded simultaneously with neural activity. For each channel, the raw activity was parsed into spiking activity (high pass filter at 250 Hz and threshold at 3.5 times the root mean square) and LFP (low pass filter at 250 Hz). Spiking Waveforms were off-line sorted using a competitive mixture decomposition method and manually refined based on waveform shape and on interspike interval using custom time amplitude window discrimination software written in MATLAB ((Kelly et al., 2007), available at https://github.com/smithlabvision/spikesort).

Receptive field mapping

The position of the visual stimuli was given by the center of the receptive field of the reference neuron. This latter was determined by hand-mapping using a bar stimulus while the animal maintained its gaze on the fixation dot. For off-line analysis, we used the minimum response field technique with a small oriented bar to map the receptive field of all the neurons across all contacts. Receptive field positions ranged from 2 to 5° eccentricity in the lower visual field, and diameter ranged from 0.5 to 1.0°.

Laminar analysis and early/late responses

For each penetration, channels were divided into superficial layers, input layer (layer 4) and deep layers, using current-source density analysis (csdplotter toolbox (https://github.com/espenhgn/CSDplotter), (Pettersen et al., 2006)) and the onset latency of the neural activity (Figure 6.b-c) (Maier et al., 2010). Early (40-110ms) and late responses (110-180ms) were measured based on the average peristimulus time histogram (PSTH) of the responses of every neurons within each layer of V1 and V2 (Figure 6.d).

Single neuron tuning selectivity

The response of each neuron to all stimuli was averaged across trials (10 repetitions). For each neuron, 3 'tuning curves' were obtained by rank-ordering the responses for each rendering type. Single neuron cue-invariance was measured using the Pearson correlation between pairs of tuning curves ($corr$ function in MATLAB). We assessed the significance of the tuning correlation based on the p-value of the Pearson correlation ($P < 0.05$). We computed the proportion of neurons that exhibit significant positive tuning correlation for pairs, double and triple pairs (i.e. all pairs) of tuning curves. The same neuron can be part of multiple proportions (for example if one neuron has significant tuning correlations for EC-AC and EC-EX correlations pairs, it will added to three sets of neurons showing significant tuning correlation, namely EC-AC, EC-EX and EC-AC and EC-EX correlation pairs). Error bars are plotted with standard error of sample proportion (SSE) (Brightwell and Dransfield, 2013):

$$SSE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The chance level of neuron proportion for each tuning correlation pairs is calculated by shuffling the tuning curve of neurons. The expected tuning correlation after would be 0 The chance level is then decided by the significance level of tuning correlation between stimulus types, which is 5% in this case. If considering only positive correlations, the chance level should then be halved to 2.5%. The chance levels of double pairs and triple pairs are then the square and the cube of the chance level of 1-pair, in this case, $6.25e-4$ and $1.5625e-5$ respectively.

The Wilson score interval (Wilson, 1927) with a 5% significance level is calculated for each fraction of neurons that show significant correlation between stimulus types. It is defined as:

$$p = \frac{p_0 + \frac{t}{2}}{1+t} \pm \frac{\sqrt{p_0(1-p_0)t + \frac{t^2}{4}}}{1+t}$$

where:

$$t = \frac{(p-\hat{p})^2}{\sigma_n^2}$$

Whether a proportion is significantly above chance is decided by whether the chance level falls into the Wilson score interval.

Z tests have been conducted to compare whether the proportions of significant neurons between different brain regions, namely p_1 and p_2, are different (Brightwell and Dransfield, 2013). With the null hypothesis:

$$H_0 : p_1 - p_2 = 0$$

The test statistic z should follow a standard normal distribution. z is defined as:

$$z = \frac{\hat{p_1} - \hat{p_2} - 0}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

where:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}, \hat{p_1} = \frac{x_1}{n_1}, \hat{p_2} = \frac{x_2}{n_2}$$

x_1, x_2 are the numbers of neurons that show significant positive correlation in the two populations, and n_1, n_2 are the total numbers of neurons in the two populations.

Procrustes transformation and Population decoding

Procrustes analysis seeks the optimal transformation, including rotation (and reflection when desirable), scaling, and translation, to align the target and comparison shapes such that the differences between the landmarks in the target shape and the transformed comparison shape are minimized (Gower, 1975). The Procrustes transformation is expressed as:

$$C^* = sCR + t$$

Consider C as a matrix with n rows and p columns in this expression, representing the comparison shape undergoing transformation. Each ith row of C corresponds to a p-dimensional landmark point. The scaling factor, denoted as s, is a scalar that modifies the size of the comparison shape. It compresses, conserves, or enlarges the shape depending on whether s is less than 1, equal to 1, or greater than 1, respectively. The rotation and reflection are embodied in R, a p-by-p transformation matrix that rotates or reflects C to align with the target shape's corresponding landmarks. t is the translation component, defined by a n-by-p matrix where each row constitutes the vector added to the points in C to translate it onto the target. Finally, C^* represents the transformed comparison shape, maintaining the same dimensions as C.

The goal is to optimize s, R, and t so that the sum of the squared differences (Euclidean distances in the case) between corresponding landmark points in the target shape (T, the same dimensions as C in our case) and transformed shape is minimized. This residual sum of the squared differences after transformation is given by:

$$S = \sum_{i=1}^{n} \|T_i - C_i^*\|^2$$

Where the i-th landmark point of the target shape and the transformed shape are represented by T_i and C_i^*, respectively. The summation runs from 1 to n, summing these squared differences for all corresponding landmark points.

In our case, Procrustes analysis was used to align the neural responses to the same set of stimuli across different rendering types (Figure 8.a). To apply the Procrustes Transformation, we obtained the mean response of each neuron across trials for every stimulus, and these responses served as the 'landmarks points' defining the shapes of the neural population. we defined the 'reference' and 'target' populations as the neural population responses to two rendering types (denoted rendering type 1 (reference population) and rendering type 2 (target population) in Figure 8). To perform the population decoding analysis (see below), We adopted a training-validation approach that was crucial to avoid any potential data leakage from the training dataset into the final calculation of the decoding accuracy. For the target population, we divided the trials into a training and a validation set. The mean response for each neuron in the training set correspond to the 'landmark points' of the target population. We applied the Procrustes Transformation to optimally align the 'shape' of the reference and of the target population. Subsequently, the parameters obtained for the Procrustes Transformation were used to transform the target population made of the trials from the validation set (denoted PT trials of type 2).

We used Support Vector Machines (SVM) to measure the decoding accuracy of the 50 stimuli for each rendering type. SVMs were trained with the complete set of trials from the reference population, and were applied to decode the target population with and without Procrustes Transformation (using 'PT trials of type 2' and 'validation trials of type 2', respectively, as denoted in Figure 8.b). This process yielded the non-PT and PT decoding accuracy.

The methodologies and parameters used in the Procrustes Transformation and the decoding processes are illustrated in Figure 8. For performing the Procrustes transformation (Figure 8.a), the reference population was obtained by averaging the responses over all 10 trials for each 50 stimuli of the rendering type 1 (rendering type 1). This process yielded a 50-by-n matrix, where n represents the number of neurons. The training set of the target population was obtained by averaging the responses of 9-of-10 trials for each 50 stimuli of the rendering type 2 (training trials of type 2), also resulting in a 50-by-n matrix. The parameters of the Procrustes Transformation were obtained y applying the transformation between the reference population and the target population made of the training trials ($procrustes$ function in MATLAB between). This transformation was then applied to the validation set of the target population made of the 1-of-10 left out trials (PT trials of type 2). In every trial sampling, we used the same leave-1-out procedure as described (yielding 50 samplings). In addition, as for each stimulus the trial index can be defined arbitrarily, we used multiple distinct trial indexes for each stimuli to increase the number of samplings of the target population (we used 15 random indexes yielding 150 samplings in total).

For PT and non-PT decoding (8.b), SVM decoders were trained to decode the 50 stimuli (MATLAB $fitcecoc$ function) using the reference population. They were then used to classify the target population made of either the validation trials (non-PT decoding) or the PT-trials of type 2 (PT decoding). During the decoding process, we ensured that the number of sampled neurons from V2 matched that of V1 as it differed for each monkey (Monkey FR: V1 - 48, V2 - 112; Monkey KO: V1 - 109, V2 - 146). For every pair of reference and target populations, the decoding accuracy was obtained with 750 samplings of the target population using the leave-1-out procedure described above. Due to the high number of samplings, standard errors (SEM) were systematically very small. For each pair of rendering types, two training/decoding procedures were possible. For example, for the AC-EC pair, training was conducted on AC (respectively, EC) followed by decoding on EC (respectively, AC). We used the average of these two estimations to obtain the final decoding accuracy (reported in Figure 3b of the main text).

Finally, we computed both an upper and a lower bound for the decoding accuracy. For the upper bound, SVMs were trained and tested on the same reference population (self-decoding) using a 10-fold cross-validation (Figure 8.c). For the lower bound, we shuffled the averaged population response of each neuron for each stimulus between the two rendering types and the same decoding procedure to obtain PT decoding accuracy was used. The decoding accuracy of the shuffled data was very close to the theoretical chance level, which is 2% for 50 classes.

Representational similarity analysis

As a second measure of cue-invariance, we used representational similarity analysis of the population response between the different rendering types. First we computed the matrix representing the similarity of the population responses for each rendering type (RSA matrix). To do so, for each rendering type, we calculated the pair-wise Pearson correlation ($corr$ function in MATLAB) between the population response vectors to all pairs of stimuli (1225 pairs for 50 stimuli). To obtain a final measure of cue-invariance, we calculated the Pearson correlation between the representation similarity

matrices for all pairs of rendering type (i.e. the correlation between the RSA matrices for the EC-AC, EC-EX and AC-EX pairs).

To measure an upper-bound on the representation similarity correlations, we first calculated the correlation between representation similarity matrices within each rendering type using the population responses made of the 2 halves of the trials (i.e. 2 sets of 5 trials out of 10 trials). The final upper-bound measure for each pair of rendering type was given by the average between their 2 upper-bound estimations.

Confidence intervals estimations were obtained by random sampling of the trials. 95% confidence interval after 1000 random samplings of trials are used to compare with the chance level of

Convolutional neural network

Implementation of Alexnet 50 and VGG 19 in MATLAB are selected for neural network analysis. The channels at the center of the feature map at each pooling layer are selected for the same analysis as the neuronal data. The input stimulus images are re-scaled to the receptive field size of the corresponding neuron and zero-padded to the size of the input of the neural networks, The outputs of the channels are treated as the firing rates of the artificial neurons.

1 Acknowledgments

We thank Karen McCracken for animal care.

References

Anzai, A., Peng, X., & Van Essen, D. C. (2007). Neurons in monkey visual area v2 encode combinations of orientations. *Nature Neuroscience*, *10*(10), 1313–1321. https://doi.org/10.1038/nn1975

Brightwell, B., & Dransfield, B. (2013). *Avoid and detect statistical malpractice: Design and analysis for biologists, with r*. InfluentialPoints.

Fitzpatrick, D. (1996). The functional organization of local circuits in visual cortex: Insights from the study of tree shrew striate cortex. *Cereb Cortex*, *6*(3), 329–41. https://doi.org/10.1093/cercor/6.3.329

Freeman, J., Ziemba, C. M., Heeger, D. J., Simoncelli, E. P., & Movshon, J. A. (2013). A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, *16*(7), 974–981. https://doi.org/10.1038/nn.3402

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science*, *315*(5814), 972–976. https://doi.org/doi:10.1126/science.1136800

Gallant, J. L., Braun, J., & Van Essen, D. C. (1993). Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science*, *259*(5091), 100–103. https://doi.org/doi:10.1126/science.8418487

Gilbert, C., Ito, M., Kapadia, M., & Westheimer, G. (2000). Interactions between attention, context and learning in primary visual cortex. *Vision Res*, *40*(10-12), 1217–26. https://doi.org/10.1016/s0042-6989(99)00234-5

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*(1), 33–51. https://doi.org/10.1007/BF02291478

Hegdé, J., & Van Essen, D. C. (2000). Selectivity for complex shapes in primate visual area v2. *J Neurosci*, *20*(5), Rc61. https://doi.org/10.1523/JNEUROSCI.20-05-j0001.2000

Hegdé, J., & Van Essen, D. C. (2007). A comparative study of shape representation in macaque visual areas v2 and v4. *Cereb Cortex*, *17*(5), 1100–16. https://doi.org/10.1093/cercor/bhl020

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/https://doi.org/10.1113/jphysiol.1968.sp008455

Kelly, R. C., Smith, M. A., Samonds, J. M., Kohn, A., Bonds, A. B., Movshon, J. A., & Sing Lee, T. (2007). Comparison of recordings from microelectrode arrays and single electrodes in the visual cortex. *The Journal of Neuroscience*, *27*(2), 261–264. https://doi.org/10.1523/jneurosci.4906-06.2007

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/https://doi.org/10.1016/j.tics.2013.06.007

Laskar, M. N. U., Sanchez Giraldo, L. G., & Schwartz, O. (2020). Deep neural networks capture texture sensitivity in v2. *Journal of Vision*, *20*(7), 21–1. https://doi.org/10.1167/jov.20.7.21

Lee, T. S., & Mumford, D. (2003). Hierarchical bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis*, *20*(7), 1434–48. https://doi.org/10.1364/josaa.20.001434

Maier, A., Adams, G. K., Aura, C., & Leopold, D. A. (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Frontiers in Systems Neuroscience*, *4*. https://doi.org/10.3389/fnsys.2010.00031

Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, *2*, 416–423 vol.2. https://doi.org/10.1109/ICCV.2001.937655

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences*, *108*(24), 9739–9746. https://doi.org/doi:10.1073/pnas.1105855108

Pettersen, K. H., Devor, A., Ulbert, I., Dale, A. M., & Einevoll, G. T. (2006). Current-source density estimation based on inversion of electrostatic forward solution: Effects of finite extent of neuronal activity and conductivity discontinuities. *Journal of Neuroscience Methods*, *154*(1-2), 116–33. https://doi.org/10.1016/j.jneumeth.2005.12.005

Sayim, B., & Cavanagh, P. (2011). What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00118

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Kar, K., Bashivan, P., Prescott-Roy, J., Schmidt, K., Yamins, D. L. K., & DiCarlo, J. J. (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv*, 407007. https://doi.org/10.1101/407007

Smith, M. A., Kelly, R. C., & Lee, T. S. (2007). Dynamics of response to perceptual pop-out stimuli in macaque v1. *Journal of Neurophysiology*, *98*(6), 3436–3449. https://doi.org/10.1152/jn.00441.2007

Tanaka, M. (2007). Recognition of pictorial representations by chimpanzees (pan troglodytes). *Animal Cognition*, *10*(2), 169–179. https://doi.org/10.1007/s10071-006-0056-1

Tang, S., Lee, T. S., Li, M., Zhang, Y., Xu, Y., Liu, F., Teo, B., & Jiang, H. (2018). Complex pattern selectivity in macaque primary visual cortex revealed by large-scale two-photon imaging. *Curr Biol*, *28*(1), 38–48.e3. https://doi.org/10.1016/j.cub.2017.11.039

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

Yan, Y., Rasch, M. J., Chen, M., Xiang, X., Huang, M., Wu, S., & Li, W. (2014). Perceptual training continuously refines neuronal population codes in primary visual cortex. *Nat Neurosci*, *17*(10), 1380–7. https://doi.org/10.1038/nn.3805

Yonas, A., & Arterberry, M. E. (1994). Infants perceive spatial structure specified by line junctions. *Perception*, *23*, 1427–1435.

Ziemba, C. M., Freeman, J., Movshon, J. A., & Simoncelli, E. P. (2016). Selectivity and tolerance for visual texture in macaque v2. *Proceedings of the National Academy of Sciences*, *113*(22), E3140–E3149. https://doi.org/10.1073/pnas.1510847113

Ziemba, C. M., Freeman, J., Simoncelli, E. P., & Movshon, J. A. (2018). Contextual modulation of sensitivity to naturalistic image structure in macaque v2. *Journal of Neurophysiology*, *120*(2), 409–420. https://doi.org/10.1152/jn.00900.2017

Ziemba, C. M., Perez, R. K., Pai, J., Kelly, J. G., Hallum, L. E., Shooner, C., & Movshon, J. A. (2019). Laminar differences in responses to naturalistic texture in macaque v1 and v2. *The Journal of Neuroscience*, *39*(49), 9748–9756. https://doi.org/10.1523/jneurosci.1743-19.2019

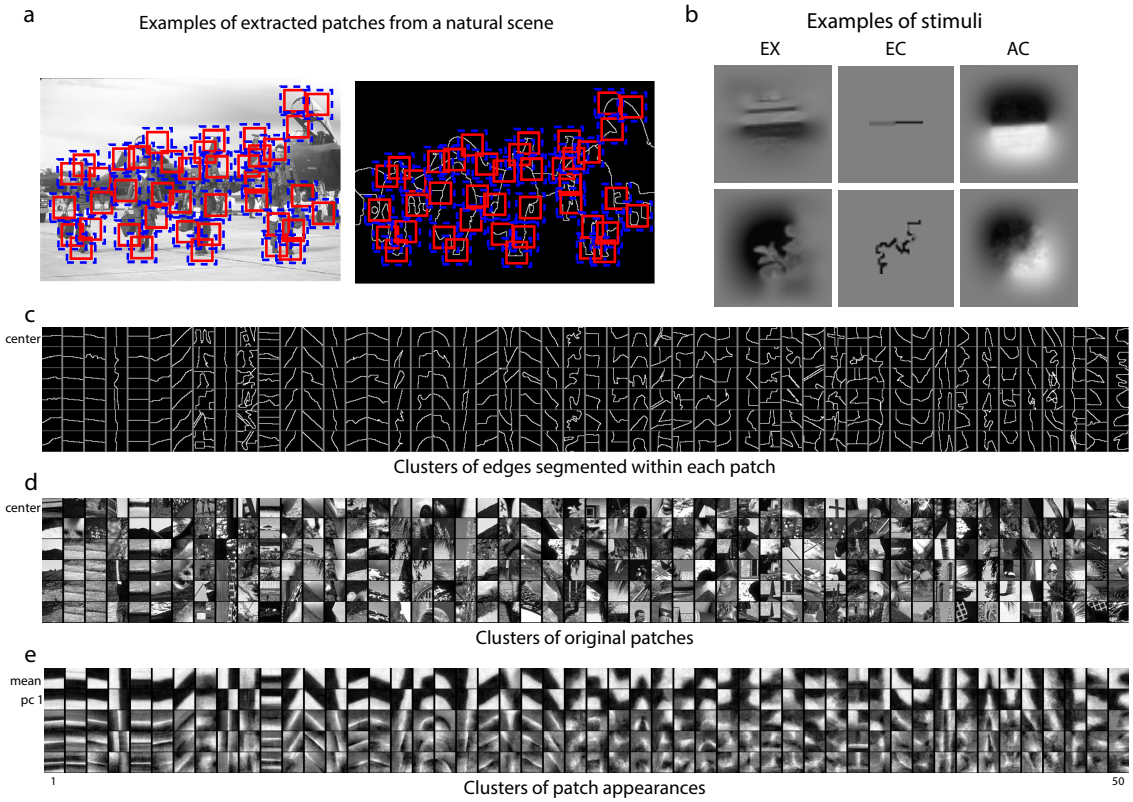

Figure 5: Contour stimuli. (a) Edge patches were extracted from natural scenes of the Berkeley Segmentation Dataset. (b) 2 example stimuli with their 3 renderings; from left to right: Example Contour stimuli (EX); Edge Contour stimuli (EC); Appearance Contour stimuli (AC). (c) Top: clusters of EC patches. Middle: clusters of EX patches associated to the edge patches in the top row. In both tables, the first row of patches corresponds to the patch at the center of the cluster. Bottom: clusters of AC patches; the patches were obtained by a PCA applied in each cluster; the first row patches corresponds to the mean appearance patch followed by 4 patches corresponding to the 4 first principle components of the PCA. In all 3 tables, the first row of patches correspond to the set of stimuli used for the visual experiment.

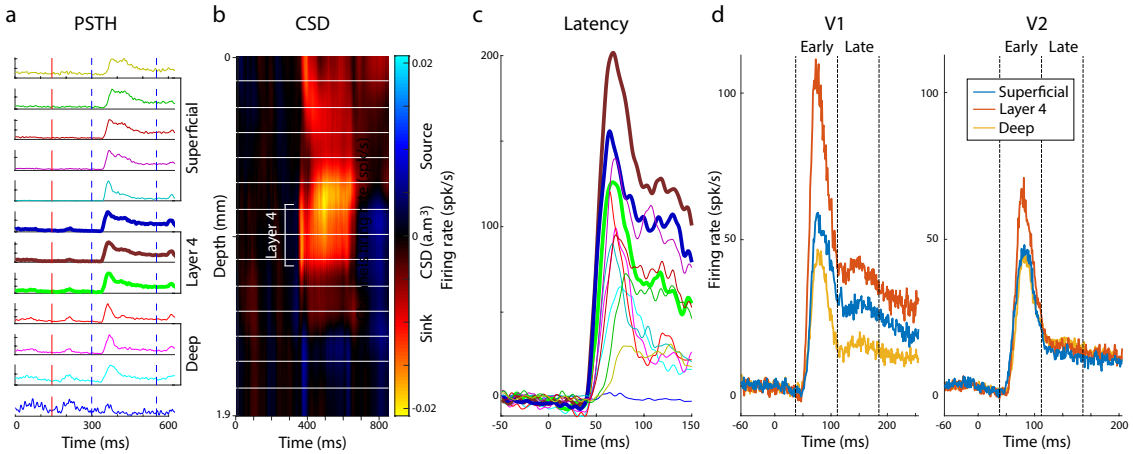

Figure 6: Laminar recording. (a) Example of recording session with the linear multi-contact probe; each row represent the PSTHs for each channel; channels are grouped into layers (deep, layer 4 and superficial) based on the CSD and onset latency information. (b) CSD associated to the example recording session. (c) All PSTHs for this session aligned on stimulus onset and used for onset latency estimation; (d) Averaged firing rate across all sessions and all neurons for V1 and V2 and grouped by their estimated layers; the vertical dashed lines indicate the windows used to define the early and late responses.

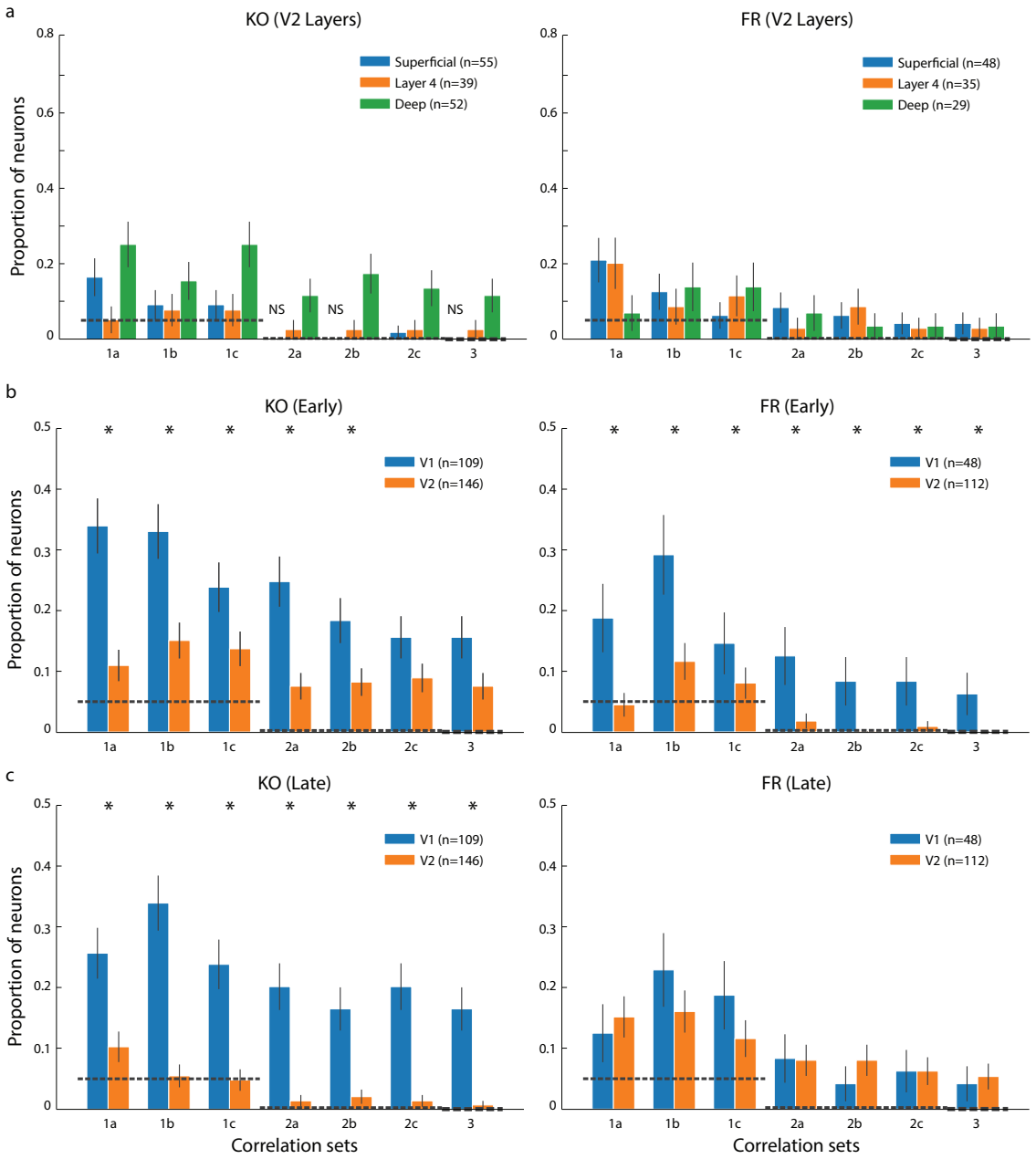

Figure 7: Tuning correlation for V1 (blue) and V2 (orange) neurons during early (top row) and late (bottom row) window for monkey KO (left) and FR (right) and the different correlations set; bar indicate the proportion of neurons that displayed significant tuning correlations for the different correlations sets; horizontal dashed lines indicate chance level for single (P=0.05), double (P=2.5×10^{-3}) and triple (P=1.25×10^{-4}) tuning correlation pairs.Tuning correlation for V1 (top row) and V2 (bottom row) neurons in superficial (blue), layer 4 (orange) and deep layers (green) for monkey KO (left) and FR (right) and the different correlations set; bar indicate the proportion of neurons that displayed significant tuning correlations for the different correlations sets; horizontal dashed line indicates chance level (p=0.05).

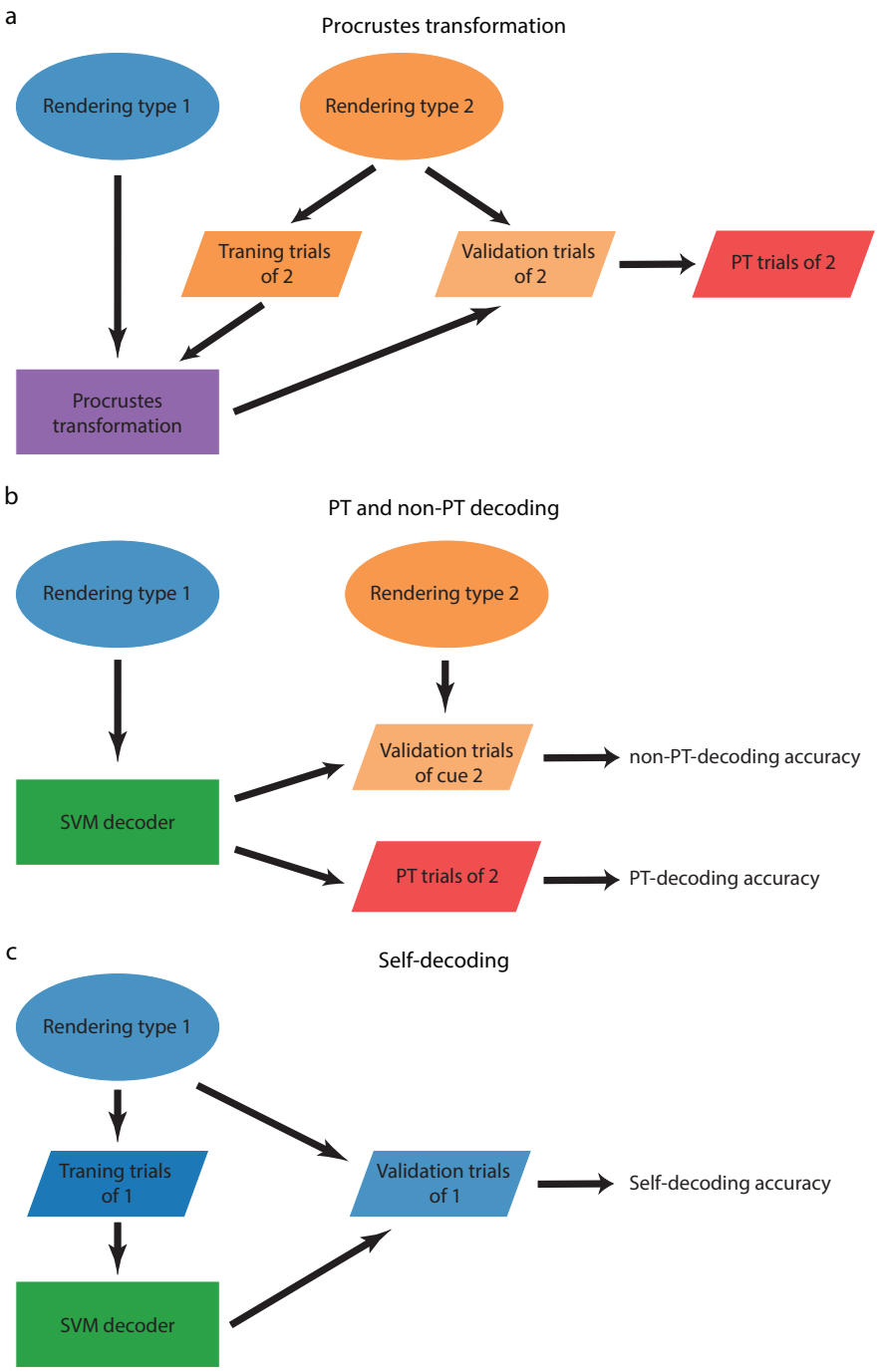

Figure 8: Decoding scheme. Each diagram represent the decoding scheme employed to estimate decoding accuracy of the population response with and with Procrustes Transformation. (a) Procrustes Transformation: generation of population response of stimuli of rendering type 2 using validation trials (population response composed of all the left over trial response for each stimuli of the second rendering type) after applying PT between stimuli of rendering type 1 and stimuli of rendering type 2 using training trials (population response of all trial responses and 9 trial responses for each stimuli of the first and second rendering type, respectively). (b) Decoding diagram for 'PT-decoding' and 'non-PT-decoding' accuracy estimation of stimuli of rendering type 2 using an SVM decoder trained on population response of stimuli of rendering type 1. (c) Decoding diagram for 'self-decoding' accuracy estimation of stimuli of rendering type 1.

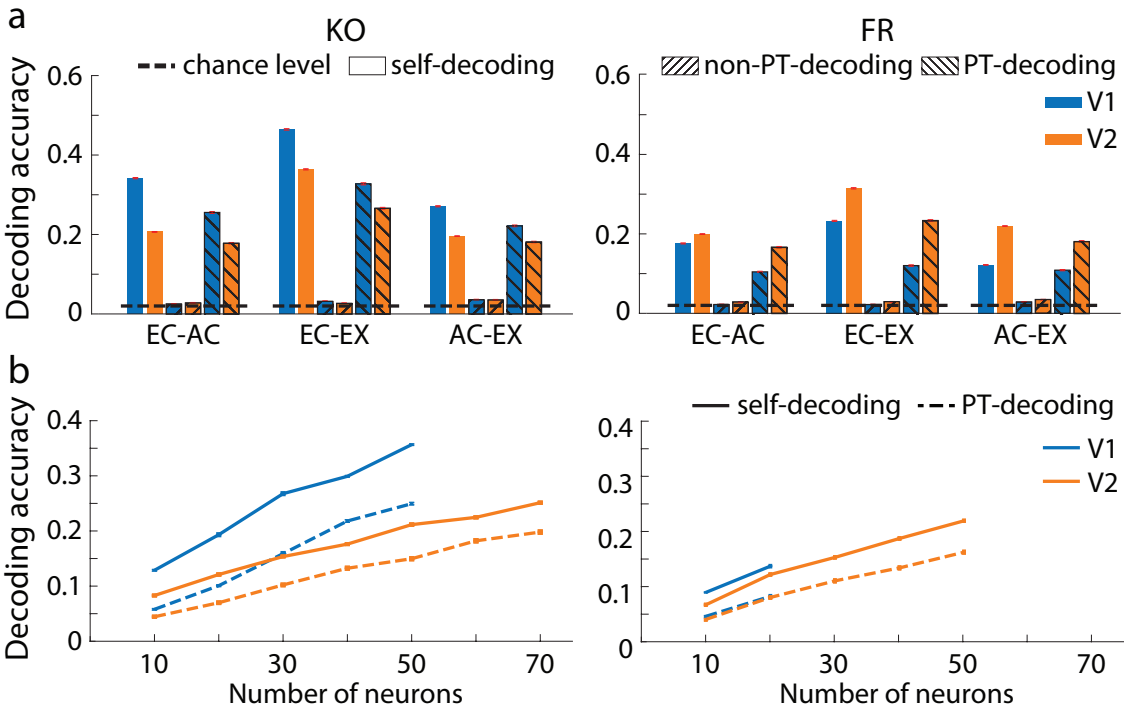

Figure 9: Population decoding with half number of neurons and effect of population size on decoding accuracy. (CHECK (Z)).

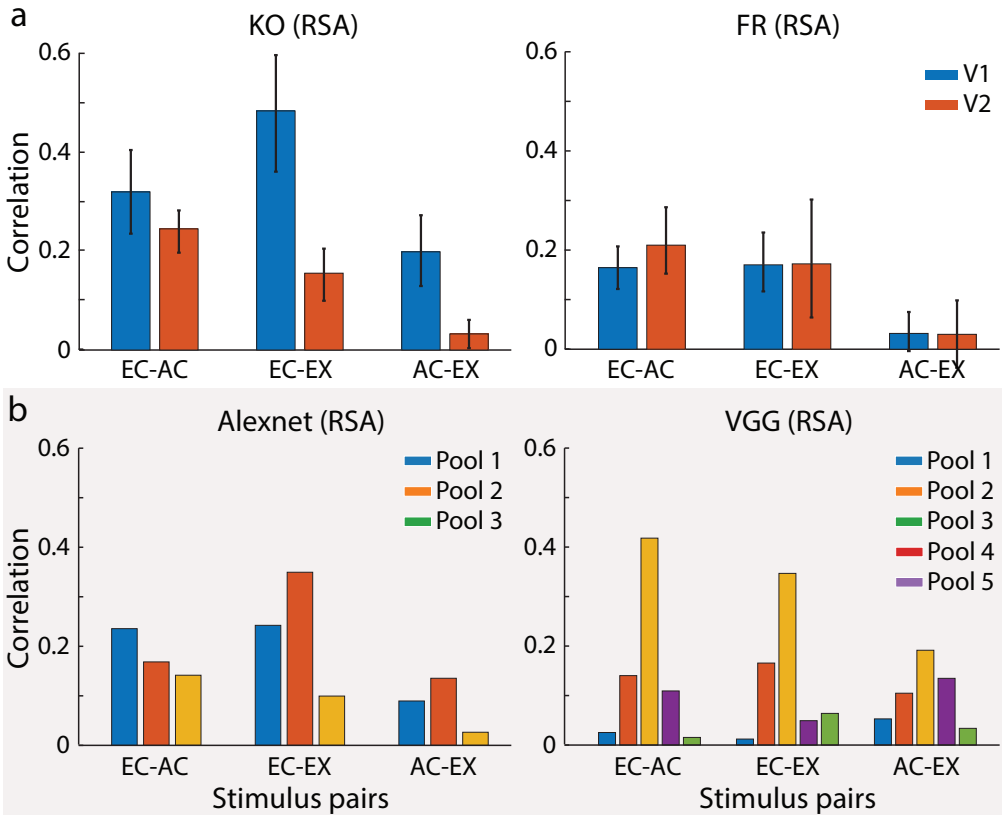

Figure 10: Representation similarity analysis (RSA). (a) Representation similarity analysis results for monkey KO and FR; x-axis indicate the pairs of rendering types used to estimate the correlation. blue and red bars indicate V1 and V2 data, respectively; second pair of bars indicate the correlation between rendering types; error bars indicate 95% (CHECK) confidence interval; first two bars (bars without borders) indicate an upper-bound on the correlation.